APPENDIX E

DEFINITIONS OF TERMS RELEVANT TO PRA AND REFERENCES FOR FURTHER READING

E.0 DEFINITIONS OF TERMS

Definitions for the specialized terms pertaining to probabilistic analysis are presented in this appendix. Some of the same terms are also defined at the beginning of each chapter, sometimes with additional examples that are relevant to concepts presented in the chapter. The definitions in this guidance are intended to be consistent with definitions used in the National Contingency Plan (NCP) and other Environmental Protection Agency (EPA) guidance, including the definitions of variability, uncertainty, and Monte Carlo simulation found in EPA's *Guiding Principles for Monte Carlo Analysis* (U.S. EPA, 1997a). Note that if a definition uses a term that is defined elsewhere in the appendix, it is highlighted in bold text.

Definitions of Terms Used in PRA	
50th percentile	The number in a distribution such that half the values in the distribution are greater than the number and half the values are less. The 50th **percentile** is equivalent to the **median**.
95th percentile	The number in a distribution such that 95% of the values in the distribution are less than or equal to the number and 5% of the values are greater than the number.
95% Upper Confidence Limit for a Mean	The 95 percent upper **confidence limit** (95% UCL) for a **mean** is defined as a value that, when repeatedly calculated for randomly drawn subsets of size *n*, equals or exceeds the true population **mean** 95% of the time. The 95% UCL provides a measure of u**ncertainty** in the **mean**; it is not a measure of **variability** and should not be confused with a 95th **percentile**. As sample size increases, the difference between the UCL for the **mean** and the true **mean** decreases, while the **95th percentile** of the distribution remains relatively unchanged, at the upper end of the distribution. EPA's Superfund program has traditionally used the 1-sided 95% UCL for the **mean** as the **concentration term** in **point estimates** of reasonable maximum exposure (**RME)** for human health **risk assessment** (U.S. EPA, 1992, 1997b).
Applicable or Relevant and Appropriate Requirements (ARARs)	Federal or state environmental standards; the NCP states that ARARs should be considered in determining **remediation goals**. ARARs may be selected as site-specific **cleanup levels**.
Arithmetic Mean (AM)	A number equal to the average value of a population or sample. Usually obtained by summing all the values in the sample and dividing by the number of values (i.e., sample size).

Definitions of Terms Used in PRA	
Assessment Endpoint	A term usually associated with **ecological risk assessment**; an explicit expression of an environmental value (ecological resource) that is to be protected, operationally defined by risk managers and risk assessors as valuable attributes of an ecological entity. Examples include 1) sustained aquatic community structure, including species composition and relative abundance and trophic structure; 2) reductions in populations of fish-eating birds; and 3) reductions in survival, reproduction or species diversity of indigenous benthic communities (U.S. EPA, 1997c, 1999a).
Backcalculation	A method of calculating a **preliminary remediation goal (PRG)** that involves algebraic rearrangement of the risk equation to solve for concentration as a function of risk, exposure, and toxicity.
Background Exposure	Exposures that are not related to the site. For example, exposure to chemicals at a different time or from locations other than the **exposure unit (EU)** of concern. Background sources may be either naturally occurring or anthropogenic (man-made).
Bayesian Analysis	Statistical analysis that describes the probability of an event as the degree of belief or confidence that a person has, given some state of knowledge, that the event will occur. Bayesian **Monte Carlo** combines a prior **probability distribution** and a **likelihood function** to yield a posterior distribution (see Appendix D for examples). Also called subjective view of probability, in contrast to the frequentist view of probability.
Bootstrap Methods	A method of sampling actual data at random, with replacement, to derive an estimate of a population **parameter** such as the **arithmetic mean** or the standard error of the **mean**. The sample size of each bootstrap sample is equal to the sample size of the original data set. Both parametric and nonparametric bootstrap methods have been developed.
Boxplot	Graphical representation showing the center and spread of a distribution, sometimes with a display of outliers (e.g., Figure 7-3). This guidance uses boxplots to represent the following **percentiles**: 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th}.
Cancer Slope Factor (CSF)	A plausible upper-bound estimate of the probability of a response per unit dose of a chemical over a lifetime. The CSF is used to estimate an upper-bound probability of an individual developing cancer as a result of a lifetime of exposure to a particular level of a potential carcinogen.
Central Limit Theorem	If random samples of size *n* are repeatedly drawn from a population of any distribution, the distribution of sample **means** converges to the normal distribution. The approximation improves as *n* increases.
Central Tendency Exposure (CTE)	A **risk descriptor** representing the average or typical individual in the population, usually considered to be the **arithmetic mean** or **median** of the risk distribution.
CTE Risk	The estimated risk corresponding to the **central tendency exposure.**

Definitions of Terms Used in PRA	
Cleanup Level	A chemical concentration chosen by the risk manager after considering both RGs and the nine selection-of-remedy criteria of the NCP (U.S. EPA, 1990; 40CFR 300.430(e)(9)(iii)). Also referred to as Final Remediation Levels (U.S. EPA, 1991), chemical-specific cleanup levels are documented in the Record of Decision (ROD). A cleanup level may differ from a **PRG** for several reasons, including various uncertainties in the risk estimate, the technical feasibility of achieving the **PRG**, and application of the nine criteria outlined in the NCP.
Coefficient of Variation	Ratio of the **standard deviation** (SD) to the **arithmetic mean** (AM) (CV=SD/AM). Dimensionless measure of the spread of a distribution, therefore, useful for comparing **probability density functions (PDFs)** for different **random variables**.
Community Advisory Group (CAG)	A group formed to provide a public forum for community members to present and discuss their needs and concerns related to the Superfund decision-making process. A CAG serves as the focal point for the exchange of information among the local community, EPA, State regulatory agency, and other pertinent Federal agencies involved in the cleanup of a Superfund site.
Community Involvement Coordinator (CIC)	As a member of the **CAG** and site team, the CIC coordinates communication plans (i.e., the **Communicty Involvement Plan (CIP)** and addresses site-specific **CAG** organizational issues.
Community Involvement Plan (CIP)	A plan that identifies community concerns and the preferences of the community for the communication of site-related issues.
Concentration Term	The concentration **variable** used in **exposure assessment**. Concentration terms are expressed in units applicable to the media of concern (e.g., mg/L for water, $\mu g/m^3$ for air; mg/kg for soil and dust.
Confidence Interval	A range of values that are likely to include a population **parameter**. **Confidence intervals** may describe a **parameter** of an input **variable** (e.g., **mean** ingestion rate) or output **variable** (e.g., **95th percentile** risk). When used to characterize **uncertainty** in a risk estimate, it is assumed that methods used to quantify **uncertainty** in the model inputs are based on statistical principles such as sampling distributions or **Bayesian** approaches. For example, given a randomly sampled data set, a 95% confidence interval for the **mean** can be estimated by deriving a sampling distribution from a Student's t distribution.
Confidence Limit	The upper or lower value of a **confidence interval**.
Continuous Variable	A **random variable** that can assume any value within an interval of real numbers (e.g., concentration).
Countably Infinite	Used to describe some discrete **random variables**, this term refers to a set of numbers that can be counted with integers (e.g., one, two, three) and that has no upper limit. Examples include the number of tosses required for a coin to show a head—we can count each toss, but it is possible that at least one more toss is needed. The number of dust particles in a volume of air is another example. Countably finite implies there is an upper limit (e.g., days of work per year).

Definitions of Terms Used in PRA	
Correlation	A quantitative relationship between two or more input **variables** of a model (e.g., body weight, inhalation rate, skin surface area). In analyses involving time-dependent **variables**, a change in one **variable** is accompanied by a change in another time-dependent, correlated **variable**. Ignoring correlations in **probabilistic risk assessment (PRA)** may lead to unrealistic combinations of values in a risk calculation. Correlations can also be defined as relationships between inputs and outputs.
Coverage	**Confidence intervals** are expected to enclose a true but unknown **parameter** according to a specified probability, such as 90% or 95%. This is the expected coverage of the **confidence interval**, given a specified significance level (alpha). The difference between the expected coverage and the actual coverage is one metric for evaluating statistical methods that yield different **confidence intervals**.
Credible Interval	A range of values that represent plausible bounds on a population **parameter**. Credible intervals may describe a **parameter** of an input **variable** (e.g., **mean** ingestion rate) or output **variable** (e.g., **95th percentile** risk). The term is introduced as an alternative to the term **confidence interval** when the methods used to quantify **uncertainty** are not based entirely on statistical principles such as sampling distributions or **Bayesian** approaches. For example, multiple estimates of an **arithmetic mean** may be available from different studies reported in the literature—using professional judgment, these estimates may support a decision to describe a range of possible values for the **arithmetic mean.**
Cumulative Distribution Function (CDF)	A graph that shows the cumulative probability of occurrence for a **random** independent **variable** (e.g., Fig. 6-1). The cumulative probability is typically given as the y-axis, ranging from 0 to 1.0. Each value *c* of the function is the probability that a random observation *x* will be less than or equal to *c*. Mathematically, the function that defines the CDF is obtained from the **PDF** by integration (in the case of a **continuous** random **variable**) or by summation (for **discrete random variables**).
Discrete Variable	A **random variable** that can assume any value within a finite set of values (e.g., number of rainfall events in one month) or at most a **countably infinite** set of values.
Empirical Distribution	A distribution obtained from actual data and possibly smoothed with interpolation techniques. Data are not fit to a particular **parametric distribution** (e.g., normal, lognormal), but are described by the **percentile** values.
Expected Value of Information (EVOI)	The expected increase in the value (or decrease in the loss) associated with obtaining more information about quantities relevant to the decision process. EVOI is a measure of the importance of **uncertainty** in risk and the potential for changing a **risk management** decision if **uncertainty** is reduced (see Appendix D).
Expert Judgment	An inferential opinion of a specialist or group of specialists within an area of their expertise. Expert judgment (alternatively referred to as professional judgment) may be based on an assessment of data, assumptions, criteria, models, and **parameters** in response to questions posed in the relevant area of expertise (see Appendix D).

Definitions of Terms Used in PRA	
Exposure Assessment	The qualitative or quantitative estimate (or measurement) of the magnitude, frequency, duration, and route of exposure. A process that integrates information on chemical fate and transport, environmental measurements, human behavior, and human physiology to estimate the average doses of chemicals received by individual receptors. For simplicity in this guidance, exposure encompasses concepts of absorbed dose (i.e., uptake and bioavailability).
Exposure Point Concentration (EPC)	The contaminant concentration within an **exposure unit** to which receptors are exposed. Estimates of the EPC represent the **concentration term** used in **exposure assessment**.
Exposure Unit (EU)	A geographic area where exposures occur to the receptor of concern during the time of interest. Receptors may be human or ecological (e.g., plants, birds, fish, mammals). For purposes of **PRA**, **probability distributions** for exposure and toxicity **variables** apply equally to all members of a population at a given exposure unit. Ecological exposure units often consider habitat and seasonality factors that enhance exposure in a spatial area usually related to home ranges.
Forward Calculations	A method of calculating a risk estimate that involves the standard arrangement of the risk equation to solve for risk as a function of concentration, exposure, and toxicity.
Frequency Distribution	A graph or plot that shows the number of observations that occur within a given interval; usually presented as a **histogram** showing the relative probabilities for each value. It conveys the range of values and the count (or proportion of the sample) that was observed across that range.
Frequentist	A term referring to classical statistics in which the probability of an event occurring is defined as the frequency of occurrence measured in an observed series of repeated trials.
Geometric Mean (GM)	The n^{th} root of the product of n observations. For lognormal distributions, the GM is equal to the median and is less than the **arithmetic mean**. For normal distributions, all three measures of central tendency (GM, **AM**, **median**) are equal.
Geostatistics	Branch of statistics that focuses on data that have a spatial or geographic components. In risk assessment, geostatistics is a general term for a variety of techniques that are typically applied to chemical concentrations in soil or groundwater in which the sampling locations are considered in quantifying the **exposure point concentration**.
Goodness-of-Fit (GoF) Test	A method for examining how well (or poorly) a sample of data can be described by a hypothesized **probability distribution** for the population. Generally involves an **hypothesis test** in which the null **hypothesis** H_0 is that a **random variable** X follows a specific **probability distribution** F_0. That is, H_0: $F = F_0$ and H_a: $F \neq F_0$.
Hazard Index (HI)	The sum of more than one **hazard quotient** for multiple substances and/or multiple exposure pathways. The HI is calculated separately for chronic, subchronic, and shorter-duration exposures.

Definitions of Terms Used in PRA	
Hazard Quotient (HQ)	The ratio of estimated site-specific exposure to a single chemical from a site over a specified period to the estimated daily exposure level, at which no adverse health effects are likely to occur.
Hazardous Substance Research Centers (HSRC)	Research centers providing free technical assistance to communities with environmental contamination programs through two distinct outreach programs: **Technical Outreach Services for Communities (TOSC)** and Technical Assistance to Brownfields Community (TAB).
High-end Risk	A **risk descriptor** representing the high-end, or upper tail of the risk distribution, usually considered to be equal to or greater than the 90^{th} **percentile**.
Histogram	A graphing technique which groups the data into intervals and displays the count of the observations within each interval. It conveys the range of values and the relative frequency (or proportion of the sample) that was observed across that range.
Hypothesis Testing	Statistical test of an assumption about a characteristic of a population. The goal of the statistical inference is to decide which of two complementary hypotheses is likely to be true.
Image Analysis	A technique in **geostatistics** used to restore a degraded image or interpret images that have been contaminated by noise or possibly some nonlinear transformation.
Independence	Two events *A* and *B* are independent if knowing whether or not *A* occurs does not change the probability that *B* occurs. Two **random variables** *X* and *Y* are independent if the joint **probability distribution** of *X* and *Y* can be expressed as the product of the individual marginal **probability distributions**. That is, $f(X, Y) = f(X) \cdot f(Y)$. Independence of *X* and *Y* is *not* synonymous with zero **correlation** (i.e., $Cor(X, Y) = 0$). If *X* and *Y* are independent, then $Cor(X, Y) = 0$; however, the converse is not necessarily true because *X* and *Y* may be related in a nonlinear fashion but still maintain zero **correlation** (Law and Kelton, 1991).
Independent and Identically Distributed (IID)	**Random variables** that are independent and have the same **probability distribution** of occurrence.
Individual-Level Effect	An **assessment endpoint** that focuses on protecting a hypothetical or real individual in a population. Individual-based models may account for unique exposure and toxicological response to chemicals among individual receptors.
Iterative Reduction (IR)	A method of calculating a **PRG** that involves successively lowering the **concentration term** until the calculated risk is acceptable. This method can be applied to any medium.
Iterative Truncation	A method of calculating a **PRG** that involves developing an expression for the **concentration term** in which high-end values are "truncated" to reduce the maximum concentration, and calculating risks associated with the reduced concentration. The method may be repeated with consecutively lower **truncation** limits until risk is acceptable. Iterative **truncation** methods avoid difficulties associated with applying **Monte Carlo analysis** to a **backcalculation**.

Definitions of Terms Used in PRA	
Kriging	A statistical interpolation method that selects the best linear unbiased estimate of the **parameter** in question. Often used as a **geostatistical** method of spatial statistics for predicting values at unobserved locations based on data from the surrounding area. Information on fate and transport of chemicals within the area lacking data can be incorporated into kriged estimates.
Kurtosis	The measure of peakedness of a distribution. A uniform distribution has a lower kurtosis than a peaked distribution such as the normal and lognormal distribution. Kurtosis is referred to as the 4th central **moment of a distribution**.
Land Method	The conventional method for calculating **uncertainty** in the **mean** concentration (e.g., **95% UCL**) when the sample data are obtained from a lognormal distribution (U.S. EPA, 1992).
Latin Hypercube Sampling (LHS)	A variant of the **Monte Carlo sampling** method that ensures selection of equal numbers of values from all segments of the distribution. LHS divides the distribution into regions of equal sampling **coverage**. Hence, the values obtained will be forced to cover the entire distribution. It is more efficient than simple random sampling, i.e., it requires fewer iterations to generate the distribution sufficiently.
Likelihood Function	A term from **Bayesian** statistics referring to a **probability distribution** that expresses the probability of observing new information given that a particular belief is true.
Local Sensitivity Analysis	Evaluation of the model sensitivity at some nominal points within the range of values of input variable(s).
Location Tag	The spatial coordinates of a sampling location (e.g., longitude, latitude).
Low-end Risk	A **risk descriptor** representing the low-end, or lower tail of the risk distribution, such as the 5th or 25th **percentile**.
Maximum Detected Concentration (MDC)	The maximum concentration detected in a sample.
Mean	**Arithmetic mean** or average; the sum of all observations divided by the number of observations. Referred to as the first central **moment of a distribution**.
Microexposure Event (MEE) Analysis	A method of assessing risk based on an aggregate sum of a receptor's contact with a contaminated medium. MEE analysis simulates lifetime exposure as the sum of many short-term, or "micro" exposures (see Appendix D). MEE approaches can be used to explore **uncertainty** associated with the model **time step** in **PRA** (e.g., use of a single value to represent a long-term average phenomenon, seasonal patterns in exposure, or intra-individual **variability**).
Mode	The most probable value of a **random variable**; a value with the largest probability or highest probability density (or mass for discrete **random variable**). The second **parameter** of a triangular distribution.

Definitions of Terms Used in PRA	
Moments of a Distribution	Similar to a **parameter**; constant that represents a mathematical description of a **random variable**. Central moments are defined with respect to the **mean**. **Mean**, **variance**, **skewness**, and **kurtosis** are the first, second, third, and fourth central moments of a **probability distribution**.
Monte Carlo Analysis (MCA) or Simulation	A technique for characterizing the **uncertainty** and **variability** in risk estimates by repeatedly sampling the **probability distributions** of the risk equation inputs and using these inputs to calculate a distribution of risk values. A set of iterations or calculations from Monte Carlo sampling is a simulation. For example, a single iteration for risk from ingestion of water may represent a hypothetical individual who drinks 2 L/day and weighs 65 kg; another iteration may represent a hypothetical individual who drinks 1 L/day and weighs 72 kg.
Monte Carlo Sampling	A method of simple random sampling used to obtain a distribution of values which may serve as an input to a **PRA**. The probability of obtaining any given sample is similar to the probability of a sample occurring within the distribution. Hence, for a given sample size, simple random sampling tends to produce values clustered around the **mean** of the distribution.
Multiple Regression Analysis	A statistical method that describes the extent, direction, and strength of the relationship between several (usually continuous) independent **variables** (e.g., exposure duration, ingestion rate) and a single continuous dependent **variable** (e.g., risk).
Nonparametric Method	A procedure for making statistical inferences without assuming that the population distribution has any specific form such as normal or lognormal. Sometimes referred to as *distribution-free* methods. Common examples are the sign test, **Spearman rank correlation**, and the **bootstrap**-t approach.
Numerical Stability	The property of a probabilistic simulation such that the a **parameter** value of the output distribution (e.g., **percentile**, **mean**, **variance**, etc.) remains sufficiently constant for a specified number of **Monte Carlo** iterations. Numerical **stability** is a measure of the precision of the output from a simulation; the tails of the distribution are typically less stable than the center. Sufficient precision is determined by professional judgment.
One-dimensional Monte Carlo Analysis (1-D MCA)	A method of simulating a distribution for an endpoint of concern as a function of **probability distributions** that characterize **variability** or **uncertainty**. In this guidance, distributions used to characterize variability may be abbreviated **PDF**v, whereas distributions used to characterize uncertainty may be abbreviated **PDF**u. It is good practice *not* to combine **PDF**s for **variability** and **uncertainty** in 1-D MCA.

Definitions of Terms Used in PRA	
Parameter	A value that characterizes the **probability distribution** of a **random variable**. For example, a normal **probability distribution** may be defined by two parameters (e.g., **AM** and SD). It is important to distinguish between this definition, and a second popular use of the term parameter when referring to an input variable in a mathematical equation or model. For this guidance, the term ***variable*** will be used to describe inputs to a model. For example, if body weight is a variable in the exposure assessment that we define with a probability distribution (e.g., normal) we would state that the variable is body weight and the parameters are the arithmetic **mean** and **standard deviation** values that characterize the normal distribution
Parametric Distribution	A theoretical distribution defined by one or more **parameters.** Examples are the normal distribution, the lognormal distribution, the triangular distribution, and the beta distribution.
Percentile	The p^{th} *percentile* of the distribution is the value such that p percent of the observations fall at or below it. Also called *quantiles* or *fractiles*; percentiles are expressed as a percent, ranging from 0 to 100, whereas quantiles or fractiles range from 0 to 1.
Point Estimate	A quantity calculated from values in a sample to represent an unknown population **parameter**. Point estimates typically represent central tendency or upper bound estimate of **variability**.
Point Estimate Risk Assessment	The familiar **risk assessment** methodology in which a single estimate of risk is calculated from a set of **point estimates**. The results provide **point estimates** of risk for the **CTE** and **RME** exposed individuals. **Variability** and **uncertainty** are discussed in a qualitative manner.
Point Pattern Analysis	A technique in **geostatistics** of restricting the analysis to location information, ignoring attribute information, addresses two location problems: (1) describing points according to spacing, and (2) describing points according to density.
Population-Level Effect	An ecological term for an **assessment endpoint** that focuses on protecting a group of individuals within a specified **exposure unit** and time that have similar exposures and toxicological responses to chemicals.
Posterior Distribution	A term from **Bayesian** statistics referring to a **probability distribution** that has been updated with new information.
Potentially Responsible Party (PRP)	Individuals, companies, or any other party that is potentially liable for Superfund cleanup costs.
Power	The probability that a test procedure detects a false null **hypothesis**; Power equals $(1-\beta)$, where β is the probability of a **Type II error** (i.e., accepting H_0 when H_a is true). Power curves are a function of a fixed significance level (α), sample size, and **variability** (**SD**).

Definitions of Terms Used in PRA	
Preliminary Remediation Goal (PRG)	A chemical concentration in an environmental medium associated with a particular exposure scenario that is expected to be protective of human health and ecosystems. PRGs may be developed based on (**ARARs**), or exposure scenarios evaluated prior to a **risk assessment** (e.g., generic PRG) or as a result of the baseline **risk assessment** (site-specific PRG). Exhibit 5-1 provides further detail on generic and site-specific PRGs.
Prior Distribution	A **Bayesian** term referring to the hypothesized, expected, or calculated **probability distribution** for an event prior to the collection of new information.
Probabilistic Risk Assessment (PRA)	A **risk assessment** that uses probabilistic methods to derive a distribution of risk or hazard based on multiple sets of values sampled for **random variables**.
Probability Density Function (PDF)	A graph that shows the probability of occurrence of an unknown or **variable** quantity. A PDF is used to characterize a **continuous random variable;** the integral of all possible values is equal to 1.0 (i.e., the area under the curve). In PRA, PDFs can be used to display the shape of the distribution for an input variable (e.g., normal distribution for ingestion rate) as well as the output from a **Monte Carlo simulation** (e.g., risk distribution).
Probability Distribution	A function that associates probabilities with the values taken by a **random variable**. A probability distribution can be displayed in a graph (e.g., **PDF** or **CDF**), summarized in a table that gives the distribution name and parameters, or expressed as a mathematical equation. In PRA, the process of selecting or fitting a distribution that characterizes **variability** or **uncertainty** can also be referred to as applying a *probability model* to characterize **variability** or **uncertainty**. In this guidance, the probability model is considered to be one source of model uncertainty.
Probability Mass Function (PMF)	A **histogram** that shows the probability of occurrence of an unknown or **variable** quantity. A PMF is used to characterize a **discrete random variable**; similar to the **PDF**, the sum of all possible values of a PMF is equal to 1.0. The mass at a point refers to the probability that the **variable** will have a value at that point.
Random Variable	A **variable** that may assume any value from a set of values according to chance. Discrete random **variables** can assume only a finite or **countably infinite** number of values (e.g., number of rainfall events per year). A random value is continuous if its set of possible values is an entire interval of numbers (e.g., quantity of rain in a year) **variable** that may assume any of a set of values. The likelihood of each value is described by a **probability distribution**.
Range Sensitivity Analysis	Evaluation of the model sensitivity across the entire range of values of the input variable(s).
Rank	If a set of values is sorted in ascending order (smallest to largest), the rank corresponds to the relative position of a number in the sequence. For example, the set {7, 5, 9, 12} when sorted gives the following sequence {5, 7, 9, 12} with ranks ranging from 1 to 4 (i.e., rank of 5 is 1, rank of 7 is 2, rank of 9 is 3, and rank of 12 is 4).

Definitions of Terms Used in PRA	
Rank Correlation (Spearman Rank Order Correlation Coefficient)	A "distribution free" or nonparametric statistic *r* that measures the strength and direction of association between the ranks of the values (not the values themselves) of two quantitative **variables**.
Remedial Investigation/Feasibility Study (RI/FS)	Studies undertaken by EPA to delineate the nature and extent of contamination, to evaluate potential risk, and to develop alternatives for cleanup.
Reasonable Maximum Exposure (RME)	The highest exposure that is reasonably expected to occur at a site (U.S. EPA, 1989, 1990). The intent of the RME is to estimate a conservative exposure case (i.e., well above the average case) that is still within the range of possible exposures.
RME Risk	The estimated risk corresponding to the **reasonable maximum exposure**.
Reference Dose (RfD)	An estimate of an exposure level for the human population, including sensitive subpopulations, that is likely to be without an appreciable risk of deleterious effects during a lifetime. Chronic RfDs are specifically developed to be protective for a long-term exposure to a chemical (e.g., >7 years) and account for **uncertainty** spanning perhaps an order of magnitude or greater.
Remediation Action Level (RAL)	Generally, a concentration such that remediation of all concentrations above this level in an **exposure unit** will result in the **95% UCL** being reduced to a level that does not pose an unacceptable risk to an individual experiencing random exposures. The RAL will depend on the **mean**, **variance**, and sample size of the concentrations within an **exposure unit** as well as considerations of acute toxicity of the chemicals of concern.
Remediation Goal	Generally, a health-based chemical concentration in an environmental medium chosen by the risk manager as appropriate for a likely land use scenario.
Risk Assessment	The use of available information to make inferences about the health effects associated with exposure of individuals or populations to hazardous materials or situations. Components of risk assessment include: hazard identification, dose-response assessment, **exposure assessment**, and **risk characterization** (NRC, 1983).
Risk Characterization	A component of **risk assessment** that describes the nature and magnitude of risk, including **uncertainty**. In assessments of Superfund sites, it includes the summary and interpretation of information gathered from previous steps in the site **risk assessment** (e.g., data evaluation, **exposure assessment**, toxicity assessment), including the results of a probabilistic analysis.
Risk Descriptor	A statistic (e.g., arithmetic **mean**, 95th **percentile**) that describes the risk to the **assessment endpoint**.
Risk Management	The process by which regulatory decisions are made using all available **risk assessment** information (including, but not limited to, the results of the **PRA**). The NCP provides nine criteria for remedial decisions (e.g., protection of human health, compliance with **ARARs**, etc.). Risk managers may include the Remedial Project Manager (RPM), section and branch chiefs, etc.

Definitions of Terms Used in PRA	
RME Range	The 90th to 99.9th **percentiles** of the risk distribution generated from a **PRA**, within which an **RME** risk value may be identified. The **95th percentile** is generally recommended as the starting point for specifying the **RME** risk in a Superfund **PRA**.
Scientific/Management Decision Point (SMDP)	A point during the **risk assessment** process when the risk assessor communicates results of the assessment at that stage to the risk manager. At this point, the risk manager determines whether the information is sufficient to arrive at a decision regarding **risk management** strategies and/or if additional information is needed to characterize risk.
Sensitivity Analysis	Process for identifying the important sources of variability and uncertainty in a model's output. Different techniques can be used in each of the 3 tiers of the tiered process for PRA (see Chapter 2). In Tier 1, sensitivity ratios are used to quantify the effects of changes in one or more model inputs on the model output. In Tiers 2 and 3, correlation analysis can be used to rank inputs based on their relative contribution to **variance** in risk. Local sensitivity refers to nominal changes in inputs within a plausible range, whereas range sensitivity refers to changes in inputs across the minimum and maximum values of the plausible range. Further explanations of the different methods for sensitivity analysis are given in Appendix A.
Sensitivity Ratio	Ratio of the change in model output per unit change in an input **variable**; also called *elasticity*.
Skewness	The measure of asymmetry of a distribution. Coefficients of skewness are zero for symmetric distributions (e.g., normal), positive for right-skewed distributions (e.g., lognormal), and negative for left-skewed distributions (e.g., specific forms of beta) . Referred to as the third central **moment of a distribution**.
Spatial Autocorrelation	The tendency of data from locations that are relatively close together to be geographically correlated.
Stakeholder	Any individual or group who has an interest in or may be affected by EPA's site decision-making process.
Stability	**Stochastic variability**, or "wobble" associated with random sampling, calculated as the average percent change in the model output after rerunning **Monte Carlo simulations** with the same set of input assumptions. Used as a metric for evaluating the adequacy of the number of iterations in a **MCA**.
Standard Deviation, Arithmetic and Geometric	Standard deviation (or arithmetic standard deviation, SD) is a common measure of the spread of a distribution. Calculated as the square root of the **variance.** The geometric standard deviation (GSD) is the anti-log of the standard deviation of the logarithms of each value. The GSD is a unitless quantity that gives a measure of the ratio of the variance to the mean, similar in concept to the **coefficient of variation**.
Step Function	A mathematical function that remains constant within each of a series of adjacent intervals but changes in value from one interval to the next. **Cumulative distribution functions** for **discrete random variables** are step functions.

Definitions of Terms Used in PRA	
Stochastic Dominance	Implies no intersection between the **CDFs**; distribution A stochastically dominates distribution B if, for every **percentile** of the **CDF**, A > B. This characteristic may not be apparent from the **PDFs** of the distributions, which may overlap.
Stochastic Process	A process involving **random variables**, and characterized by **variability** in space or time.
Target Population	The set of all receptors that are potentially at risk. Sometimes referred to as the "population of concern". A sample population is selected for statistical sampling in order to make inferences regarding the target population (see Appendix B, Section B.3.1, Concepts of Populations and Sampling).
Technical Assistance Grant (TAG)	A federal grant that is intended to provide a community with the opportunity to hire independent experts to help evaluate and explain the results of a **risk assessment**
Technical Outreach Services for Communities (TOSC)	A service of the **HSRC** with the aim to provide independent technical information and assistance to help communities with hazardous substance pollution problems.
Thiessen (Voronoi) Polygon Analysis	A method of spatial statistics in which an area is subdivided into subregions, or polygons, in order to predict values at unobserved locations.
Time Step	A **variable** in all exposure models that refers to the unit of time for which a **random value** is considered representative of intra-individual **variability** (e.g., average daily ingestion rates for an individual from one year to the next). A time step may be equal to an entire exposure duration (e.g., 30 years), or a fraction of the exposure duration during which changes in input **variables** may be expected (e.g., one year). Time steps need not be identical for all exposure **variables**, and should address the most rapidly changing **variable** in the risk equation. Time step can be an important consideration for **MEE analysis**.
Toxicity Reference Value (TRV)	A numerical expression of a chemical's dose-response relationship that is used in **ecological risk assessment**.
True Mean Concentration	The actual average concentration in an **exposure unit**. Even with extensive sampling, the true **mean** cannot be known. Only an estimate of the true **mean** is possible. A greater number of representative samples increases confidence that the estimate of the **mean** more closely represents the true **mean**.
Truncation	The process of setting lower and upper limits on the range of a distribution, in order to avoid unrealistic values for exposure **variables** (e.g., > 100% bioavailability). Most often used for continuous, unbounded **probability distributions** (e.g., normal).

Definitions of Terms Used in PRA	
Two-dimensional Monte Carlo Analysis (2-D MCA)	An advanced modeling technique that uses two stages of random sampling, also called nested loops, to distinguish between **variability** and **uncertainty** in exposure and toxicity **variables**. The first stage, often called the inner loop, involves a complete **1-D MCA** simulation of **variability** in risk. In the second stage, often called the outer loop, **parameters** of the **probability distributions** are redefined to reflect **uncertainty**. These loops are repeated many times resulting in multiple risk distributions, from which **confidence intervals** are calculated to represent **uncertainty** in the population distribution of risk.
Type I Errors	False positive; the error made when the null **hypothesis** is rejected in favor of the alternative, when in fact the null **hypothesis** is true.
Type II Errors	False negative; the error made when the null **hypothesis** is accepted when in fact the alternative **hypothesis** is true.
Uncertainty	Lack of knowledge about specific **variables**, **parameters**, models, or other factors. Examples include limited data regarding the concentration of a contaminant in an environmental medium and lack of information on local fish consumption practices. Uncertainty may be reduced through further study.
Variability	True heterogeneity or diversity in characteristics among members of a population (i.e., inter-individual variability) or for one individual over time (intra-individual variability). For example, body weights of a study population at one point in time will exhibit variability, and body weight will change as an individual ages. Further study (e.g., increasing sample size, n) will not reduce variability, but it can provide greater confidence in quantitative characterizations of variability.
Variable	A quantity that can assume many values.
Variance	Measure of the spread of a distribution, equal to the square of the **standard deviation** (SD). Calculated as the average of the squares of the deviations of the observations from their **mean**. Variance is referred to as the second central **moment of a distribution**.
Z-score	The value of a normally distributed **random variable** that has been standardized to have a **mean** of zero and a SD of one by the transformation $Z=(X-\mu)/\sigma$. Statistical tables typically give the area to the left of the z-score value. For example, the area to the left of z =1.645 is 0.95. Z-scores indicate the direction (+/-) and number of **standard deviations** away from the **mean** that a particular datum lies assuming X is normally distributed. Microsoft Excel's *NORMSDIST(z)* function gives the probability p such that $p=Pr(Z \leq z)$, while the *NORMSINV(p)* function gives the z-score z_p associated with probability p such that $p=Pr(Z \leq z_p)$.

E.1.0 ADDITIONAL INFORMATION

Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis (Morgan and Henrion, 1990) and *Probabilistic Techniques in Exposure Assessment* (Cullen and Frey, 1999) provide excellent philosophical and practical treatises on probabilistic risk assessment. These works are highly recommended to risk assessors who wish to know more about probabilistic risk assessment. The *Summary Report for the Workshop on Monte Carlo Analysis* (U.S. EPA, 1996) and the *Summary Report for the Workshop on Selecting Input Distributions for Probabilistic Assessments* (U.S. EPA, 1999b) are other sources of information to learn more about PRA. Other additional references for reading are listed in this Appendix.

REFERENCES FOR APPENDIX E

Cullen, A.C. and H.C. Frey. 1999. *Probabilistic Techniques in Exposure Assessment.* A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press.

Law, A.M. and W.D. Kelton. 1991. Simulation Modeling and Analysis. McGraw-Hill, Inc., NY.

Morgan, G.M. and M. Henrion, 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, NY.

National Research Council (NRC). 1983. *Risk Assessment in the Federal Government: Managing the Process*. National Academy Press. Washington, DC.

U.S. EPA. 1989. *Risk Assessment Guidance for Superfund* (RAGS): *Volume I. Human Health Evaluation Manual* (HHEM) *(Part A, Baseline Risk Assessment).* Interim Final. Office of Emergency and Remedial Response, Washington, DC. EPA/540/1-89/002. NTIS PB90-155581.

U.S. EPA. 1990. National Oil and Hazardous Substances Pollution Contingency Plan. Final Rule. 40 CFR 300: *55 Federal Register*, 8666-8865, March 8.

U.S. EPA. 1991. *Risk Assessment Guidance for Superfund* (RAGS)*, Volume I: Human Health Evaluation Manual* (HHEM)*, Part B, Development of Risk-Based Preliminary Remediation Goals.* Office of Emergency and Remedial Response, Washington, DC. EPA/540/R-92/003. NTIS PB92-963333.

U.S. EPA. 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. Office of Solid Waste and Emergency Response, Washington, DC. OSWER Directive No. 9285.7-081.

U.S. EPA. 1997a. *Memorandum from Deputy Administrator Fred Hansen on the Use of Probabilistic Techniques (including Monte Carlo Analysis) in Risk Assessment, and Guiding Principles for Monte Carlo Analysis*. Office of Research and Development, Washington, DC. EPA/630/R-97/001. May 15.

U.S. EPA. 1997b. *Lognormal Distribution in Environmental Applications*. Office of Research and Development, and Office of Solid Waste and Emergency Response, Washington, DC. EPA/600/R-97/006. December.

U.S. EPA. 1997c. *Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments.* Interim Final. U.S. Environmental Protection Agency, Environmental Response Team (Edison, NJ). June 5.

U.S. EPA. 1999a. Issuance of Final Guidance: Ecological Risk Assessment and Risk Management Principles for Superfund Sites. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response. OSWER Directive 9285.7-28 P. Stephen D. Luftig for Larry D. Reed. October 7.

U.S. EPA. 1999b. *Summary Report for the Workshop on Selecting Input Distributions for Probabilistic Risk Assessment.* Risk Assessment Forum. EPA/630/R-98/004.

REFERENCES FOR FURTHER READING

Baird, BF. 1989. *Managerial Decisions Under Uncertainty*. John Wiley & Sons, Inc., NY.

Bevington, P.R. 1969. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, NY.

Bratley, P., B.L. Fox, and L.E. Schrage. 1987. *A Guide to Simulation.* Springer-Verlag, NY.

Burmaster, D.E. and P.D. Anderson. 1994. Principles of Good Practice for the Use of Monte Carlo Techniques in Human Health and Ecological Risk Assessment. *Risk Anal.* 14(4):477–481.

Clemen, R, 1990. *Making Hard Decisions.* Duxbury Press.

Conover, W.I. 1971. *Practical Nonparametric Statistics*. John Wiley & Sons, NY.

Cox, D.C. and P. Baybutt. 1981. *Methods for Uncertainty Analysis: A Comparative Survey*. *Risk Anal* 1(4):251–258.

Cullen, A.C. and H.C. Frey. 1999. *Probabilistic Techniques in Exposure Assessment.* Plenum Press, NY.

D'Agostino, R. and M.A. Stephens, eds. 1986. *Goodness-of-Fit Techniques*. Marcel Dekker, Inc., NY.

Devroye, L. 1986. *Non-Uniform Random Deviate Generation.* Springer-Verlag, NY.

Evans, M., N. Hastings, and B. Peacock. 1993. *Statistical Distributions*. John Wiley & Sons, NY.

Finkel, A.M. 1990. *Confronting Uncertainty in Risk Management: A Guide for Decision-Makers.* Resources for the Future, Washington, DC.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* Van Nostrand Reinhold, NY.

Hamby, D.M. 1994. A review of Techniques for Parameter Sensitivity Analysis of Environmental Models. *Environ. Monit. and Assess*. 32:135–154.

Hammersley, J.M. and D.C. Handscomb. 1964. *Monte Carlo Methods*. John Wiley & Sons, NY.

Hertz, D.B. and H. Thomas. 1983. *Risk Analysis and Its Applications*. John Wiley & Sons, NY.

Hertz, D.B. and H. Thomas. 1984. *Practical Risk Analysis - An Approach Through Case Studies*. John Wiley & Sons, NY.

Hoffman, F.O. and J.S. Hammonds. 1992. *An Introductory Guide to Uncertainty Analysis in Environmental and Health Risk Assessment.* ES/ER/TM-35, Martin Marietta.

Hoffman, F.O. and J.S. Hammonds. 1994. *Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability. Risk Anal* 14(5):707–712.

Iman, R.L. and W.J. Conover. 1982. A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables. *Commun. Stat*, Part B 11:311–331.

Iman, R.L. and J.C. Helton. 1988. An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Anal*. 8(1):71–90.

Iman, R.L., J.M. Davenport, and D.K. Zeigler. 1980. *Latin Hypercube Sampling (A Program Users Guide)*. Technical Report SAND 79:1473, Sandia Laboratories, Albuquerque.

Johnson, M.E. 1987. *Multivariate Statistical Simulation*. John Wiley & Sons, NY.

Johnson, N.L. and S. Kotz. 1970. *Continuous Univariate Distributions*. Vols. 1 & 2. John Wiley & Sons, NY.

Johnson, N.L., S. Kotz, and A.W. Kemp. 1992. *Univariate Discrete Distributions*. John Wiley & Sons, NY.

Kendall, M. and A. Stuart. 1979. *Advanced Theory of Statistics, Volume I - Distribution Theory, Volume II - Inference and Relationship*. MacMillan, Inc., NY.

Kennedy, W.J. and E. Gentle. 1980. *Statistical Computing*. Marcel Dekker, Inc., NY.

LePage, R. and L. Billard. 1992. *Exploring the Limits of Bootstrap*. John Wiley & Sons, NY.

Lipton, J., W.D. Shaw, J. Holmes, and A. Patterson. 1995. Short Communication: Selecting Input Distributions for use in Monte Carlo Analysis. *Regul. Toxicol. Pharmacol*. 21:192–198.

McKone, T.E. and K.T. Bogen. 1992. Uncertainties in Health Risk Assessment: An Integrated Case Based on Tetrachloroethylene in California Groundwater. *Regul. Toxicol. Pharmacol.* 15:86–103.

Megill, R.E., ed. 1985. *Evaluating and Managing Risk*. Penn Well Books, Tulsa, OK.

Morgan, G.M. and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, NY.

NCRP. 1996. Commentary No. 14. *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination*. National Committee on Radiation Programs, Scientific Committee 64-17. Washington, DC.

Palisade Corporation. 1994. *Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 1-2-3*. Windows Version Release 3.0 User's Guide, Palisade Corporation, Newfield, NY.

Press, W.H., B.P. Flannery, S.A. Teulolsky, and W.T. Vetterling. 1989. *Numerical Recipes in Pascal: the Art of Scientific Computing*. Cambridge University Press, NY.

Press, W.H., S.A. Teulolsky, W.T. Vetterling, and B.P. Flannery. 1992. *Numerical Recipes in FORTRAN: the Art of Scientific Computing*. Cambridge University Press, NY.

Press, W.H., S.A. Teulolsky, W.T. Vetterling, and B.P. Flannery. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, NY.

Read, T. and N. Cressie. 1988. *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer-Verlag, NY.

Rohatgi, V.K. 1984. *Statistical Inference*. John Wiley & Sons, NY.

Rubenstein, R.Y. 1981. *Simulation and the Monte Carlo Method*. John Wiley & Sons, NY.

Sachs, L. 1984. *Applied Statistics - A Handbook of Techniques*. Springer-Verlag, NY.

Saltelli, A and J. Marivort. 1990. Non-Parametric Statistics in Sensitivity Analysis for Model Output: A Comparison of Selected Techniques. *Reliab. Eng. Syst. Saf*. 28:299–253.

Schneider, H. 1986. *Truncated and Censored Distributions from Normal Populations*. Marcel Dekker, Inc., NY.

Seiler, F.A. 1987. Error Propagation for Large Errors. *Risk Anal*. 7(4):509–518.

Seiler, F.A. and J.L. Alvarez. 1996. On the Selection of Distributions for Stochastic Variables. *Risk Anal*. 16(1):5–18.

Slob, W. 1994. Uncertainty Analysis in Multiplicative Models. *Risk Anal*. 14(4):571–576.

Smith, A.E., P.B. Ryan, and J.S. Evans. 1992. The Effect of Neglecting Correlations when Propagating Uncertainty and Estimating the Population Distribution of Risk. *Risk Anal*. 12(4):467474.

Smith, R.L. 1994. Uses of Monte Carlo Simulation for Human Exposure Assessment at a Superfund Site. *Risk Anal* 14(4):433–439.

Sokal, R. and R. Rohlf. 1981. *Biometry: The Principles and Practice of Statistics in Biological Research*. Second Edition. W.H. Freeman & Co., NY.

U.S. EPA. 1978. *Source Assessment: Analysis of Uncertainty - Principles and Applications*. EPA/600/2-79-004.

U.S. EPA. 1992a. *Guidelines for Exposure Assessment*. *Federal Register*. 57(104):22888-22938. May 29.

U.S. EPA. 1992b. *Guidelines for Carcinogenic Risk Assessment. Federal Register*. 51(185):33992-34003. May 29.

U.S. EPA. 1996. *Summary Report for the Workshop on Monte Carlo Analysis*. Office of Research and Development, Washington, DC. EPA/630/R-96/010.

U.S. EPA. 1999. *Guidelines for Carcinogenic Risk Assessment*. Review Draft. Risk Assessment Forum. Washington, DC. NCEA-F-0644.

U.S. EPA. 2001. *Guidelines for Carcinogenic Risk Assessment. Federal Register.* 66(230):59593-59594. November 29.

Wilks, D.S. 1995. *Statistical Methods in the Atmospheric Sciences, An Introduction*. Academic Press, San Diego.